This page will be used to document the discussions and progress of the UAF Technical Team. It will be organized around the questions or issues that the team is wrestling with and will be updated periodically based on the team's meetings, e-mails, and discussions. Team members are encouraged to add questions, issues, and comments at any time.

Presentation highlighting the current status of the UAF Grid project as of January 14, 2010.

Contents

UAF Grid Questions and Issues:

What criteria should be used to determine which datasets are "valid" and should be included?

Some of the criteria for determining whether a dataset is "valid" are a matter of policy. (e.g. Are non-NOAA datasets "valid" in UAF? What about a non-NOAA dataset that is proxied through a NOAA server?) Policies change over time. Also there are valid reasons to have minor policy differences in different part of the system. (e.g. For select groups of users the discovery system may choose to expose metadata about datasets that are not accessible.) So we don't want the policy choices embedded deeply into our system. We need to decide on a strategy to ensure this is the case.

- A Web service into the metadata catalog that returns all information relevant the "validity" of a dataset?
- Provide a "clean" THREDDS catalog that excludes the datasets from the raw THREDDS UAF catalog that are deemed to be invalid
- other ideas??

What additional datasets should be added to the catalog?

The table below contains a list of submitted datasets that are not currently included in the UAF Grid Catalog.

Server Organization	Server URL	Description of Datasets served	Why includ UA Cata
CO-OPS	http://opendap.co-ops.nos.noaa.gov/thredds/catalog.xml	Forecast models	need aggrega
NODC	http://data.nodc.noaa.gov/thredds/catalog.xml	Various satellite datasets, gridded in situ analysis fields, and more	invalid dataset aggrega
NOMADS		Forecast models	need aggrega

OceanNomads			
NCEP RTOFS	http://edac-dap2.northerngulfinstitute.org/thredds/catalog/rtofs3d/catalog.xml	Forecast models	need aggrega
	http://edac-dap2.northerngulfinstitute.org/thredds/catalog/rtofs/catalog.xml		
NOAA Coastwatch	http://coastwatch.noaa.gov/thredds/catalog.xml	Satellite data - Chlorophyll-a Concentration and Reflectance	some n aggrega but ther aggrega that sho in the c
NDBC	http://dods.ndbc.noaa.gov/thredds/catalog/data/catalog.xml	Observational data: TAO/OceanSITES	not grid data
IOOS Regions		Various model runs as well as in-situ data	not grid data an needing aggrega
AOOS	http://137.229.40.88/opendap/catalog.xml	Various model runs as well as in-situ data	not grid data an needing aggrega
NANOOS	http://agate.coas.oregonstate.edu:8080/thredds/catalog.xml	Oregon Coastal Ocean Simulator	not grid data an needing aggrega
CENCOOS	http://cencoos.org:8080/thredds/catalog.xml	Various model runs as well as in-situ data	not grid data an needing aggrega
SCCOOS	http://ourocean.jpl.nasa.gov:8080/thredds/catalog.xml	SCB Forecast-LAS Grid, PWS Forecast Data, Latest SCB Forecast--ROMS Grid, Pacific Data, Global G1SST, HYCOM Data, SCB Nowcast/Forecast Aggregation Catalog, OSSE Models Ensemble, MARCOOS SST Catalog, Solomon Data	not grid data an needing aggrega
PACIOOS	http://oos.soest.hawaii.edu/thredds/catalog.xml	HiOOS THREDDS Data Server	not grid data an needing aggrega

GLOS	http://michigan.glin.net:8080/thredds/catalog.xml	GLOS THREDDS Server	not gric data an needing aggrega
NERACOOS	http://rocky.umeoce.maine.edu:8080/thredds/catalog.xml	University of Maine School of Marine Sciences Ocean Modeling Group	not gric data an needing aggrega
	http://coast-enviro.er.usgs.gov/thredds/catalog.xml	Adriatic Sea Project, Bathymetry, Carolina Coastal Change Project, Gulf of Maine Interoperability Demo, IOOS Top Level THREDDS Data Server Catalogs, IOOS Clean Catalog (only aggregated, CF-compliant data), Massachusetts Bay Project, ROMS Forecast Models Catalog, Vineyard Sound/Middle Ground Project, Island Wake Simulations, MVCO ROMS-SWAN Model Runs, Latte Runs, PV 2002 Runs, CMGP Oceanographic Time Series Data, Buzzards Bay Runs	not gric data an needing aggrega
	http://geoport.whoi.edu:8081/thredds/catalog.xml	USGS SedTrans THREDDS Server	not gric data an needing aggrega
	http://blackburn.whoi.edu:8080/thredds/catalog.xml	Various model runs as well as in-situ data	not gric data an needing aggrega
	http://www.smast.umassd.edu:8080/thredds/catalog.xml	All Model Input and Output files, FVCOM NECOFS Forecasts	not gric data an needing aggrega

What additional datasets should be added to the catalog?

	http://www.jcoot.unh.edu/thredds/catalog.xml	MM5 Model Runs, WRF Forecast Model, UNH Water Balance Model, UNH Water Balance Models	not gric data an needing aggrega
MACOORA	http://colossus.dl.stevens-tech.edu:8080/thredds/catalog.xml	VTHREDDS Catalog: The New York Harbor Observing and Prediction System (NYHOPS)	not gric data an needing aggrega
	http://tashtego.marine.rutgers.edu:8080/thredds/catalog.xml	IMCS Catalog	not gric data an needing aggrega
	http://aqua.smast.umassd.edu:8080/thredds/catalog.xml	Oceanographic Modeling and Analysis Laboratoty THREDDS catalogs	not gric data an needing aggrega
SECOORA	http://omglnx1.meas.ncsu.edu:8080/thredds/catalog.xml	GOMTOX (Gulf of Maine) Ocean Model, SABGOM (South Atlantic Bight and Gulf of Mexico) Ocean Model, SABGOM (South Atlantic Bight and Gulf of Mexico) 2007 hindcast, MABGOM (Middle Atlantic Bight and Gulf of Maine) Ocean Model	not gric data an needing aggrega
CARICOOS	http://dm1.caricoos.org/thredds/catalog.xml	Chlorophyll, NWS Wind Forecast	not gric data an needing aggrega
GCOOS	http://csanady.tamu.edu:8080/thredds/catalog.xml	TGLO/TABS Model documentation, ROMS History Files, NCEP Surface Winds on Regular	not gric data an needing aggrega

		Grid, ROMS Surface Currents on Regular Grid, ROMS Surface Currents on Curvilinear Grid, ROMS Bottom Currents on Curvilinear Grid, ROMS 3-D Currents on Curvilinear Grid, NCEP Surface Winds on Regular Grid, ROMS Surface Currents on Regular Grid	

How should data discovery be accomplished?

See UAF Metadata.

How should data and dataset aggregation be handled?

Aggregation is vital to the usability of gridded data, because it can turn a collection of thousands of individual files (one per time step) into a single multi-dimensional gridded dataset. Some of the data that will be accessible through the THREDDS catalog will be collections of non-aggregated datasets. Sometimes files will be aggregated, yet the individual files will also be exposed through THREDDS -- i.e. the same data are presented in two different ways. How should we handle these non-aggregated datasets? (Part of the solution should be a way to encourage the data provider to aggregate her data and coaching on how to do so at low effort.)

- Proposed short-term approach:

1. identify them and aggregate them manually for now
2. communicate the aggregation ncML back to the data provider
3. keep track of what works for communications

- Proposed long term approach:

1. if a dataset (a virtual netCDF-CF file) ==> send back scripts for selected applications
2. if a "catalog" (a branching point in the THREDDS tree) ==> send back THREDDS catalog reference

- Steve Hankin and John Caron discussed this topic on 11/24. Here is a quick summary:
  - The challenge of aggregations is simultaneously dealing with changeable file collections, high performance, configuration simplicity, and robustness.
  - An improved aggregation capability is planned for TDS 4.2 (V4.1 is not yet out). It is planned to include remote scans
  - John expects to start working on TDS 4.2 around Jan. 1-ish.
  - Steve volunteered the UAF-grid group to be Alpha testers of the capabilities
  - It would be helpful to John if we would provide short write ups of the most important use cases.

How often should the catalog be updated?

Some datasets will have a real-time component to them -- say, updated hourly or daily. What is the process for determining the update interval for a given dataset? Should we engineer a way to share the assessment regarding this characteristic about a dataset?